Probability of Error in Information-Hiding Protocols

Konstantinos Chatzikokolakis, Catuscia Palamidessi, Prakash Panangaden

Probability of Error in Information-Hiding Protocols*

Konstantinos Chatzikokolakis Catuscia Palamidessi
INRIA and LIX, École Polytechnique
Palaiseau, France
{kostas,catuscia}@lix.polytechnique.fr

Prakash Panangaden
McGill University
Montreal, Quebec, Canada
prakash@cs.mcgill.ca

Abstract

Randomized protocols for hiding private information can often be regarded as noisy channels in the information-theoretic sense, and the inference of the concealed information can be regarded as a hypothesis-testing problem. We consider the Bayesian approach to the problem, and investigate the probability of error associated to the inference when the MAP (Maximum Aposteriori Probability) decision rule is adopted. Our main result is a constructive characterization of a convex base of the probability of error, which allows us to compute its maximum value (over all possible inputs' distribution), and to identify functional upper bounds for it. As a side result, we are able to substantially improve the Hellman-Raviv and the Santhi-Vardy bounds expressed in terms of conditional entropy. We then discuss an application of our methodology to the Crowds protocol, and in particular we show how to compute the bounds on the probability that an adversary break anonymity.

1 Introduction

Information-hiding protocols try to hide the relation between certain facts, that we wish to maintain hidden, and the *observable* consequences of these facts. Example of such protocols are the anonymity protocols like Crowds [21], Onion Routing [26], and Freenet [7]. Often these protocols use randomization to obfuscate the link between the hidden information and the observed events. Crowds, for instance, tries to conceal the identity of the originator of a message by forwarding randomly the message till its destination, so that if an attacker intercepts the message, it cannot be sure whether the sender is the originator or just a forwarder.

In most cases, protocols like the above can be regarded as information-theoretic channels, where the inputs are the facts to keep hidden, the outputs are the observables, and the matrix represents the correlation between the facts and the observed events, in terms of conditional probabilities. An adversary can try to infer the facts from the observed events with the Bayesian method, which is based on the principle of assuming an a priori probability distribution on the hidden facts (*hypotheses*), and deriving from that (and from the matrix) the a posteriori distribution after a certain event has been observed. It is well known that the best strategy for the adversary is to apply the MAP (Maximum Aposteriori Probability) criterion, which, as the name says, dictates to choose the hypothesis with the maximum a posteriori probability. "Best" means that this criterion induces the smallest probability of guessing the wrong hypothesis. The probability of error, in this case, is also called *Bayes' risk*.

Even if the adversary does not know the a priori distribution, the method is still valid asymptotically, under the condition that the matrix' rows are all pairwise distinguished. By repeating the experiment, in fact, the contribution of the a priori probability becomes less and less relevant for the computation of the a posteriori probability, and it "washes out" in the limit. Furthermore, the probability of error converges to 0 in the limit [8]. If the rows are all equal, namely if the channel has capacity 0, then the Bayes' risk is maximal and does not converge to 0. This is the ideal situation, from the point of view of information-hiding protocols. In practice, however, it is difficult to achieve such degree of privacy. We are then interested in maximizing the Bayes' risk, so to make the convergence to 0 as slow as possible. The main purpose of this paper is to investigate the Bayes' risk, in relation to the channel's matrix, and its bounds.

There are many bounds known in literature for the Bayes' risk. One of these is the *equivocation bound*, due to Rényi [22], which states that the probability of error is bound by the conditional entropy of the channel's input given the output. Later, Hellman and Raviv improved this bound by half [13]. Recently, Santhi and Vardy have proposed a new bound, that depends exponentially on the (opposite of the) conditional entropy, and which considerably improves the Hellman-Raviv bound in the case of multi-

*This work has been partially supported by the INRIA DREI Équipe Associée PRINTEMPS. The work of Konstantinos Chatzikokolakis and Catuscia Palamidessi has been also supported by the INRIA ARC project ProNoBiS.

hypothesis testing [23].

1.1 Contribution

The main contributions of this paper are the following:

1. We consider what we call "the corner points" of a piecewise linear function, and we propose criteria to compute the maximum of the function, and to identify concave upper bounds for it, based on the analysis of its corner points only.

2. We consider the hypothesis testing problem in relation to an information-theoretic channel. In this context, we show that the probability of error associated to the MAP rule is piecewise linear, and we give a constructive characterization of a set of corner points, which turns out to be finite. Together with the previous results, this leads to constructive methods to compute the maximum probability of error over all the channel's input distributions, and to define tight functional upper bounds.

3. As a side result of the above study, we are able to improve on the Hellman-Raviv and the Santhi-Vardy bounds, which express relations between the Bayes risk and the conditional entropy. The Santhi-Vardy bound, which is better than the Hellman-Raviv one when we consider more than two hypotheses, is tight (i.e. it coincides with the Bayes' risk) on the corner points only in the case of channels with capacity 0. Our improved bound is tight on those points for every channel. The same holds with respect to the Hellman-Raviv bound (the latter is better than the Santhi-Vardy one in the case of two hypotheses).

4. We show how to apply the above results to randomized protocols for information hiding. In particular, we work out in detail the application to Crowds, and derive the maximum probability of error for an adversary who tries to break anonymity, and bounds on this probability in terms of conditional entropy, for any input distribution.

1.2 Related work

Probabilistic notions of anonymity and information-hiding have been explored in [4, 12, 1, 2]. We discuss the relation with these works in detail in Section 5.

A recent line of work has been dedicated to exploring the concept of anonymity from an information-theoretic point of view [24, 10]. The main difference with our approach is that in those works the anonymity degree is expressed in terms of input entropy, rather than conditional entropy. More precisely, the emphasis is on the lack of information of the attacker about the distribution of the users, rather than on the capability of the protocol to conceal this information despite of the observables that are made available to the attacker. Moreover, a uniform user distribution is assumed, while in this paper we abstract from the user distribution in the functional sense.

In [17, 18] the ability to have covert communication as a result of non-perfect anonymity is explored. Those works focus on the possibility of constructing covert channels by the users of the protocol, using the protocol mechanisms, and on measuring the amount of information that can be transferred through these channels. In [18] the authors also suggest that the channel's capacity can be used as an asymptotic measure of the worst-case information leakage. Another information-theoretical approach is the one of [9], where the authors use the notion of *relative entropy* to define the degree of anonymity.

In the field of information flow and non-interference there is a line of research which is related to ours. There have been various works [16, 11, 5, 6, 14] in which the the *high information* and the *low information* are seen as the input and output respectively of a channel. From an abstract point of view, the setting is very similar; technically it does not matter what kind of information we are trying to conceal, what is relevant for the analysis is only the probabilistic relation between the input and the output information. The conceptual and technical novelties of this paper w.r.t. the above works are explained in Section 1.1. We believe that our results are applicable more or less directly also to the field of non-interference.

The connection between the adversary's goal of inferring a secret from the observables, and the field of "hypothesis testing", has been explored in other papers in literature, see in particular [15, 19, 20, 3]. To our knowledge, however, [3] is the only work exploring the Bayes' risk in connection to the channel associated to an information-hiding protocol. More precisely, [3] considers a framework in which anonymity protocols are interpreted as particular kinds of channels, and the degree of anonymity provided by the protocol as the converse of the channel's capacity (an idea already suggested in [18]). Then, [3] considers a scenario in which the adversary can enforce the re-execution of the protocol with the same input, and studies the Bayes' risk on the repeated experiment. The focus is on how the adversary can approximate the MAP rule when the a priori distribution is not known, and the main result of [3] on this topic is the investigation of the characteristics of the matrix that make this task possible or impossible. In the present paper, on the contrary, we study the Bayes' risk as a function of the a priori distribution, and we give criteria to compute tight bounds for it.

1.3 Plan of the paper

Next section recalls some basic notions in information theory, and about hypothesis testing and probability of error. Section 3 proposes some methods to identify tight bounds for a function that is generated by a set of corner points. Section 4 presents the main result of our work, namely a constructive characterization of the corner points of Bayes' risk. In Section 5 we discuss the relation with some probabilistic information-hiding notions in literature. Finally, Section 6 illustrates an application of our results to the anonymity protocol Crowds.

The report version of this paper, containing the proofs, is available on line at: http://www.lix.polytechnique.fr/~catuscia/papers/ProbabilityError/full.pdf

2 Information theory, hypothesis testing and probability of error

In this section we briefly revise some basic notions in information theory and hypothesis testing that will be used trough the paper. We refer to [8] for more details.

A *channel* is a tuple $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$ where $\mathcal{A}, \mathcal{O}$ are the sets of input and output values respectively and $p(o|a)$ is the conditional probability of observing output $o \in \mathcal{O}$ when $a \in \mathcal{A}$ is the input. In this paper, we assume that both $\mathcal{A}$ and $\mathcal{O}$ are finite with cardinality n and m respectively. We will also sometime use indices to represent their elements: $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$ and $\mathcal{O} = \{o_1, a_2, \ldots, o_m\}$. The $p(o|a)$'s constitute what is called the *matrix* of the channels. The usual convention is to arrange the a's by rows and the o's by columns.

In general, we consider the input of a channel as *hidden information*, and the output as *observable information*. The set of input values can also be regarded as a set of *mutually exclusive* (hidden) *facts* or *hypotheses*. A probability distribution $p(\cdot)$ over $\mathcal{A}$ is called *a priori probability*, and it induces a probability distribution over $\mathcal{O}$ (called *marginal probability* of $\mathcal{O}$). In fact

$$p(o) = \sum_a p(a, o) = \sum_a p(o|a)\, p(a)$$

where $p(a, o)$ represents the joint probability of a and o, and we use its Bayesian definition $p(a, o) = p(o|a)p(a)$.

When we observe an output o, the probability that the corresponding input has been a certain a is given by the conditional probability $p(a|o)$, also called *a posteriori probability* of a given o, which in general is different from $p(a)$. This difference can be interpreted as the fact that observing o gives us evidence that changes our degree of belief in the hypothesis a. The a priori and the a posteriori probabilities of a are related by Bayes' theorem:

$$p(a|o) = \frac{p(o|a)\, p(a)}{p(o)}$$

In hypothesis testing we try to infer the *true* hypothesis (i.e. the input fact that really took place) from the observed output. In general it is not possible to determine the right hypothesis with certainty. We are interested, then, in minimizing the *probability of error*, i.e. the probability of making the wrong guess. Formally, the probability of error is defined as follows. Given the *decision function* $f : \mathcal{O} \to \mathcal{A}$ adopted by the observer to infer the hypothesis, let $E_f : \mathcal{A} \to \mathcal{O}$ be the function that gives the *error region* of f when $a \in \mathcal{A}$ has occurred, namely:

$$E_f(a) = \{o \in \mathcal{O} \mid f(o) \neq a\}$$

Let $\eta_f : \mathcal{A} \to [0, 1]$ be the function that associates to each $a \in \mathcal{A}$ the probability that f gives the the wrong input fact when $a \in \mathcal{A}$ has occurred, namely:

$$\eta_f(a) = \sum_{o \in E_f(a)} p(o|a)$$

The probability of error for f is then obtained as the sum of the probability of error for each possible input, averaged on the probability of the input:

$$P_f = \sum_a p(a)\, \eta_n(a)$$

In the Bayesian framework, the best possible decision function f_B, namely the decision function that minimizes the probability of error, is obtained by applying the MAP (*Maximum Aposteriori Probability*) criterion, that chooses an input a with a maximal $p(a|o)$. Formally:

$$f_B(o) = a \;\; \Rightarrow \;\; \forall a' \;\; p(a|o) \geq p(a'|o)$$

The probability of error associated to f_B, aka *Bayes' risk*, is then given by

$$P_e = 1 - \sum_o p(o)\, \max_a p(a|o) = 1 - \sum_o \max_a p(o|a)\, p(a)$$

Note that f_B, and the Bayes' risk, depend on the inputs' a priori probability. The input distributions can be represented as the elements $\vec{x} = (x_1, x_2, \ldots, x_n)$ of a domain $D^{(n)}$ defined as

$$D^{(n)} = \{\vec{x} \mid \sum_i x_i = 1 \text{ and } \forall i \; x_i \geq 0\}$$

where the correspondence is given by $\forall i \; x_i = p(a_i)$. In the rest of the paper we will assume the MAP rule and view the Bayes' risk as a function $P_e : D^{(n)} \to [0, 1]$ defined by

$$P_e(\vec{x}) = 1 - \sum_i \max_j p(o_i|a_j) x_j \qquad (1)$$

There are some notable results in literature relating the Bayes' risk to the information-theoretic notion of *conditional entropy*, aka *equivocation*. Let us first recall the concept of *random variable* and its *entropy*. A random variable A is determined by a set of values $\mathcal{A}$ and a probability distribution $p(a)$ over $\mathcal{A}$. The entropy of A, $H(A)$, is given by

$$H(A) = -\sum_a p(a) \log p(a)$$

The entropy measures the uncertainty of a random variable. It takes its maximum value $\log n$ when A's distribution is uniform and its minimum value 0 when A is constant. We usually consider the logarithm with a base 2 and measure entropy in *bits*.

Now let A, O be random variables. The *conditional entropy* $H(A|O)$ is defined as

$$H(A|O) = -\sum_o p(o) \sum_a p(a|o) \log p(a|o)$$

The conditional entropy measures the amount of uncertainty of A when O is known. It can be shown that $0 \leq H(A|O) \leq H(A)$. It takes its maximum value $H(A)$ when O reveals no information about A, i.e. when A and O are independent, and its minimum value 0 when O completely determines the value of A.

Comparing $H(A)$ and $H(A|O)$ gives us the concept of *mutual information* $I(A;O)$, which is defined as

$$I(A;O) = H(A) - H(A|O)$$

Mutual information measures the amount of information that one random variable contains about another random variable. In other words, it measures the amount of uncertainty about A that we lose when observing O. It can be shown that it is symmetric ($I(A;O) = I(A;O)$) and that $0 \leq I(A;O) \leq H(A)$. The maximum mutual information between A and O over all possible input distributions $p(a)$ is known as the channel's *capacity*:

$$C = \max_{p(a)} I(A;O)$$

The capacity of a channel gives the maximum rate at which information can be transmitted using this channel.

Given a channel, let $\vec{x}$ be the a priori distribution on the inputs. Recall that $\vec{x}$ also determines a probability distribution on the outputs. Let A and O be the random variables associated to the inputs and outputs respectively. The Bayes' risk is related to $H(A|O)$ by the Hellman and Raviv's bound [13]:

$$P_e(\vec{x}) \leq \frac{1}{2} H(A|O) \tag{2}$$

and by the Santhi and Vardy's bound [23]:

$$P_e(\vec{x}) \leq 1 - 2^{-H(A|O)} \tag{3}$$

We remark that, while the bound (2) is tighter than (3) in case of binary hypothesis testing, i.e. when $n = 2$, (3) gives a much better bound when n becomes larger. In particular the bound in (3) is always limited by 1, which is not the case for (2).

3 Convexly generated functions and their bounds

In this section we characterize a special class of functions on probability distributions, and we present various results regarding their bounds which lead to methods to compute their maximum, to prove that a concave function is an upper bound, and to derive an upper bound from a concave function. The interest of this study is that the probability of error will turn out to be a function in this class.

We start by recalling some basic notions: let $\mathbb{R}$ be the set of real numbers. The elements $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ constitute a set of *convex coefficients* iff $\forall i \ \lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. Given a vector space V, a *convex combination* of $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k \in V$ is any vector of the form $\sum_i \lambda_i \vec{x}^i$ where the λ_i's are convex coefficients. A subset S of V is *convex* iff every convex combination of vectors in S is still in S. It is easy to see that for any n the domain $D^{(n)}$ of probability distributions of dimension n is convex. Given a subset S of V, the *convex hull* of S, which we will denote by $ch(S)$, is the smallest convex set containing S. It is well known that $ch(S)$ always exists.

We now introduce (with a slight abuse of terminology) the concept of *convex base*:

Definition 3.1 *Given the vector sets S, U, we say that U is a convex base for S iff $U \subseteq S$ and $S \subseteq ch(U)$.*

In the following, given a vector $\vec{x} = (x_1, x_2, \ldots, x_n)$, we will use the notation $(\vec{x}, f(\vec{x}))$ to denote the vector (in a space with an additional dimension) $(x_1, x_2, \ldots, x_n, f(\vec{x}))$. Similarly, given a vector set S in a n-dimensional space, we will use the notation $(S, f(S))$ to represent the vector set $\{(\vec{x}, f(\vec{x})) \mid \vec{x} \in S\}$ in a $(n+1)$-dimensional space. The notation $f(S)$ represents the image of f in S, i.e. $f(S) = \{f(\vec{x}) \mid \vec{x} \in S\}$.

We are now ready to introduce the class of functions that we announced at the beginning of this section:

Definition 3.2 *Given a vector set S, a convex base U of S, and a function $f : S \to \mathbb{R}$, we say that $(U, f(U))$ is a set of corner points of f iff $(U, f(U))$ is a convex base for $(S, f(S))$. We also say that f is convexly generated by $f(U)$*[1].

[1] To be more precise we should say that f is convexly generated by $(U, f(U))$.

Of particular interest are the functions that are convexly generated by a finite number of corner points. This is true for *piecewise linear functions* in which S can be decomposed into finitely many convex polytopes (n-dimensional polygons) and f is equal to a linear function on each of them. Such functions are convexly generately by the (finite) set of vertices of these polytopes.

We now give a criterion for computing the maximum of a convexly generated function.

Proposition 3.3 *Let $f : S \to \mathbb{R}$ be convexly generated by $f(U)$. If $f(U)$ has a maximum element b, then b is the maximum value of f on S.*

Proof Let b be the maximum of $f(U)$. Then for every $u \in U$ we have that $f(u) \leq b$. Consider now a vector $\vec{x} \in S$. Since f is convexly generated by $f(U)$, there exist $\vec{u}^1$, $\vec{u}^2$, ..., $\vec{u}^k$ in U such that $f(\vec{x})$ is obtained by convex combination from $f(\vec{u}^1)$, $f(\vec{u}^2)$, ..., $f(\vec{u}^k)$ via some convex coefficients λ_1, λ_2, ..., λ_k. Hence:

$$\begin{array}{rcll} f(\vec{x}) & = & \sum_i \lambda_i f(\vec{u}^i) & \\ & \leq & \sum_i \lambda_i b & \text{since } f(\vec{u}^i) \leq b \\ & = & b & \lambda_i\text{'s being convex combinators} \end{array}$$

□

Note that if U is finite then $f(U)$ has always a maximum element.

Next, we propose a method for proving (functional) upper bounds for f, when they are in the form of *concave* functions.

We recall that, given a vector set S, a function $g : S \to \mathbb{R}$ is concave iff for any $\vec{x}^1, \vec{x}^2, \ldots, \vec{x}^k \in S$ and any set of convex coefficients $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ we have

$$\sum_i \lambda_i \, g(\vec{x}^i) \leq g(\sum_i \lambda_i \vec{x}^i)$$

Proposition 3.4 *Let $f : S \to \mathbb{R}$ be convexly generated by $f(U)$ and let $g : S \to \mathbb{R}$ be concave. Assume that for all $\vec{u} \in U$ $f(\vec{u}) \leq g(\vec{u})$ holds. Then we have that g is an upper bound for f, i.e.*

$$\forall \vec{x} \in S \;\; f(\vec{x}) \leq g(\vec{x})$$

Proof Let $\vec{x}$ be an element of S. Since f is convexly generated, there exist $\vec{u}^1$, $\vec{u}^2$, ..., $\vec{u}^k$ in U such that $(\vec{x}, f(\vec{x}))$ is obtained by convex combination from $(\vec{u}^1, f(\vec{u}^1))$, $(\vec{u}^2, f(\vec{u}^2))$, ..., $(\vec{u}^k, f(\vec{u}^k))$ via some convex coefficients λ_1, λ_2, ..., λ_k. Hence:

$$\begin{array}{rcll} f(\vec{x}) & = & \sum_i \lambda_i f(\vec{u}^i) & \\ & \leq & \sum_i \lambda_i g(\vec{u}^i) & \text{since } f(\vec{u}^i) \leq g(\vec{u}^i) \\ & \leq & g(\sum_i \lambda_i \vec{u}^i) & \text{by the concavity of } g \\ & = & g(\vec{x}) & \end{array}$$

□

Finally, we give a method to obtain tight functional upper bounds from concave functions.

Proposition 3.5 *Let $f : S \to \mathbb{R}$ be convexly generated by $f(U)$ and let $g : S \to \mathbb{R}$ be concave. Assume that for each $\vec{u} \in U$ if $g(\vec{u}) = 0$ then $f(\vec{u}) \leq 0$. Consider the set $R = \{f(\vec{u})/g(\vec{u}) \mid \vec{u} \in U, g(\vec{u}) \neq 0\}$. If R has a maximum element c, then the function $c\,g$ is a tight functional upper bound for f, i.e.*

$$\forall \vec{x} \in S \;\; f(\vec{x}) \leq c\, g(\vec{x})$$

and f and $c\,g$ coincide at least in one point.

Proof Since c is the maximum of R, we have that, for every $\vec{u} \in U$ with $g(\vec{u}) \neq 0$, $f(\vec{u}) \leq c\, g(\vec{u})$ holds. On the other hand, if $g(\vec{u}) = 0$, then $f(\vec{u}) \leq 0 = c\, g(\vec{u})$. Hence by Proposition 3.4 we have that $c\,g$ is an upper bound for f. Furthermore, if $\vec{v}$ is the vector for which $f(\vec{u})/g(\vec{u})$ is maximum, then $f(\vec{v}) = c\, g(\vec{v})$ so the bound is tight. □

Note that, if U is finite, then the maximum element of R always exists.

3.1 An alternative proof for the Hellman-Raviv and Santhi-Vardy bounds

Using Proposition 3.4 we can give an alternative, simpler proof for the bounds in (2) and (3). We start with the following proposition, whose proof can be found in the appendix:

Proposition 3.6 *Let $f : D^{(n)} \to \mathbb{R}$ be the function $f(\vec{y}) = 1 - \max_j y_j$. Then f is convexly generated by $f(U)$ with $U = U_1 \cup U_2 \cup \ldots \cup U_n$ where, for each k, U_k is the set of all vectors that have value $1/k$ in exactly k components, and 0 everywhere else.*

Consider now the functions $g, h : D^{(n)} \to \mathbb{R}$ defined as

$$g(\vec{y}) = \frac{1}{2} H(\vec{y}) \quad \text{and} \quad h(\vec{y}) = 1 - 2^{-H(\vec{y})}$$

where (with a slight abuse of notation) H represents the entropy of the distribution $\vec{y}$, i.e. $H(\vec{y}) = -\sum_j y_j \log y_j$.

We have that both g and h satisfy the conditions of Proposition 3.4 with respect to f, and therefore

$$\forall \vec{y} \in D^{(n)} \;\; f(\vec{y}) \leq g(\vec{y}) \;\text{ and }\; f(\vec{y}) \leq h(\vec{y}) \qquad (4)$$

The rest of the proof proceeds as in [13] and [23]: Let $\vec{x}$ represent an a priori distribution on $\mathcal{A}$ and let the above $\vec{y}$ denote the a posteriori probabilities on $\mathcal{A}$ with respect to a certain observable o, i.e. $y_j = p(a_j|o) =$

$(p(o|a_j)/p(o))\,x_j$. Then $P_e(\vec{x}) = \sum_o p(o) f(\vec{y})$, so from (4) we obtain

$$P_e(\vec{x}) \leq \sum_o p(o)\frac{1}{2}H(\vec{y}) = \frac{1}{2}H(A|O) \tag{5}$$

and

$$P_e(\vec{x}) \leq \sum_o p(o)(1 - 2^{-H(\vec{y})}) \leq 1 - 2^{-H(A|O)} \tag{6}$$

where the last step in (6) is obtained by applying Jensen's inequality. This concludes the alternative proof of (2) and (3).

We end this section with two remarks. First, we note that g coincides with f only on the points of U_1 and U_2, whereas h coincides with f on all U. In fact, if $\vec{u}^k$ is an element of U_k, we have $f(\vec{u}^1) = 0 = 1/2 \log 1 = g(\vec{u}^1)$, $f(\vec{u}^2) = 1/2 = 1/2 \log 2 = g(\vec{u}^2)$, and for $k \geq 2$, $f(\vec{u}^k) = 1 - 1/k < 1$ while $g(\vec{u}^k) = 1/2 \log k > 1$. On the other hand, for all k, $h(\vec{u}^k) = 1 - 2^{-\log k} = f(\vec{u}^k)$. This explains, intuitively, why (3) is a better bound than (2) for dimensions higher than 2.

Second, we observe that, although h is a very tight bound for f, when we average h and f on the output probabilities to obtain $\sum_o p(o)(1 - 2^{-H(\vec{y})})$ and $P_e(\vec{x})$ respectively, and then we apply Jensen's inequality, we usually loosen this bound a lot, as we will see in some examples later. The only case in which we do not loose anything is when the channel has capacity 0 (maximally noisy channel), i.e. all the rows of the matrix are the same. In the general case of non-zero capacity, however, this implies that if we want to obtain a better bound we need to follow a different strategy. In particular, we need to find directly the corner points of P_e instead than those of the f defined above. This is what we are going to do in the next section.

4 The corner points of the Bayes' risk

In this section we present our main contribution, namely we show that P_e is convexly generated by $P_e(U)$ for a finite U, and we give a constructive characterization of U, so that we can apply the results of previous section to compute tight bounds on P_e.

The idea behind the construction of such U is the following: recall that the Bayes' risk is given by $P_e(\vec{x}) = 1 - \sum_i \max_j p(o_i|a_j)x_j$. Intuitively, this function is linear as long as, for each i, the j which gives the maximum $p(o_i|a_j)x_j$ remains the same while we vary $\vec{x}$. When, for some i and k, the maximum becomes $p(o_i|a_k)x_k$, the function changes its inclination and then it becomes linear again. The exact point in which the inclination changes is a solution of the equation $p(o_i|a_j)x_j = p(o_i|a_k)x_k$. This equation actually represents a hyperplane (a space in $n-1$ dimensions, where n is the cardinality of $\mathcal{A}$) and the inclination of P_e changes in all its points for which $p(o_i|a_j)x_j$ is maximum, i.e. it satisfies the inequation $p(o_i|a_j)x_j \geq p(o_i|a_\ell)x_\ell$ for each ℓ. The intersection of $n-1$ hyperplanes of this kind, and of the one determined by the equation $\sum_j x_j = 1$, is a vertex $\vec{v}$ such that $(\vec{v}, P_e(\vec{v}))$ is a corner point of P_e.

Definition 4.1 *Given a channel $\mathcal{C} = \langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, the family $\mathbb{S}(\mathcal{C})$ of the systems generated by $\mathcal{C}$ is the set of all systems of inequations of the following form:*

$$\begin{array}{rcl}
p(o_{i_1}|a_{j_1})x_{j_1} & = & p(o_{i_1}|a_{j_2})x_{j_2} \\
p(o_{i_2}|a_{j_2})x_{j_2} & = & p(o_{i_2}|a_{j_3})x_{j_3} \\
 & \vdots & \\
p(o_{i_k}|a_{j_{r-1}})x_{j_{r-1}} & = & p(o_{i_k}|a_{j_r})x_{j_r} \\
x_j & = & 0 \quad \text{for } j \notin \{j_1, j_2, \ldots, j_r\} \\
x_1 + x_2 + \ldots + x_n & = & 1 \\
p(o_{i_h}|a_{j_h})x_{j_h} & \geq & p(o_{i_h}|a_\ell)x_\ell \quad \text{for } 1 \leq h, \ell \leq r, n
\end{array}$$

where n is the cardinality of $\mathcal{A}$, $r \leq n$, and $j_1, j_2, \ldots, j_r$ are pairwise different.

A system is called *solvable* if it has solutions. Note that a system of the kind considered in the above definition has at most one solution.

We are now ready to state our main result:

Theorem 4.2 *Given a channel $\mathcal{C}$, the Bayes' risk P_e associated to $\mathcal{C}$ is convexly generated by $P_e(U)$, where U is constituted by the solutions to all solvable systems in $\mathbb{S}(\mathcal{C})$.*

Proof We need to prove that, for every $\vec{v} \in D^{(n)}$, there exist $\vec{u}^1, \vec{u}^2, \ldots, \vec{u}^t \in U$, and convex combinators $\lambda_1, \lambda_2, \ldots, \lambda_t$ such that

$$\vec{v} = \sum_i \lambda_i \vec{u}^i \quad \text{and} \quad P_e(\vec{v}) = \sum_i \lambda_i P_e(\vec{u}^i)$$

Let us consider a particular $\vec{v} \in D^{(n)}$. In the following, for each i, we will use j_i to denote the index j for which $p(o_i|a_j)v_j$ is maximum. Hence, we can rewrite $P_e(\vec{v})$ as

$$P_e(\vec{v}) = \sum_i p(o_i|a_{j_i})v_{j_i} \tag{7}$$

We proceed by induction on n.

Base case ($n = 2$) In this case U is the set of solutions of all the systems of the form

$$\{p(o_i|a_1)x_1 = p(o_i|a_2)x_2 \ , \ x_1 + x_2 = 1\}$$

and $\vec{v} \in D^{(2)}$. Let c be the minimum between v_1 and the minimum $x \geq 0$ such that

$$p(o_i|a_1)(v_1 - x) = p(o_i|a_2)(v_2 + x) \quad \text{for some } i$$

Analogously, let d be the minimum between v_2 and the minimum $x \geq 0$ such that

$$p(o_i|a_2)(v_2 - x) = p(o_i|a_1)(v_1 + x) \quad \text{for some } i$$

Let us define $\vec{v}^1$, $\vec{v}^2$ as

$$\vec{v}^1 = (v_1 - c, v_2 + c) \qquad \vec{v}^2 = (v_1 + d, v_2 - d)$$

Consider the convex coefficients

$$\lambda_1 = \frac{d}{c+d} \qquad \lambda_2 = \frac{c}{c+d}$$

A simple calculation shows that

$$\vec{v} = \lambda_1 \vec{v}^1 + \lambda_2 \vec{v}^2$$

It remains to prove that

$$P_e(\vec{v}) = \lambda_1 P_e(\vec{u}^1) + \lambda_2 P_e(\vec{u}^2) \tag{8}$$

To this end, we need to show that P_e is defined in $\vec{v}^1$ and $\vec{v}^2$ by the same formula as (7), i.e. that for each i and $k \neq j_i$ the inequation $p(o_i|a_{j_i})v^1_{j_i} \geq p(o_i|a_k)v^1_k$ holds, and similarly for $\vec{v}^2$.

Let i and k be given. If $j_i = 1$, and consequently $k = 2$, we have that for some $x \geq 0$ the equality $p(o_i|a_2)(v_2 - x) = p(o_i|a_1)(v_1 + x)$ holds. Therefore:

$$\begin{array}{rcll} p(o_i|a_1)v^1_1 & = & p(o_i|a_1)(v_1 - c) & \text{by definition of } \vec{v}^1 \\ & = & p(o_i|a_1)(v_1 - x) & \text{since } c \leq x \\ & = & p(o_i|a_2)(v_2 + x) & \text{by definition of } x \\ & \geq & p(o_i|a_2)(v_2 + c) & \text{since } c \leq x \\ & = & p(o_i|a_1)v^1_2 & \text{by definition of } \vec{v}^1 \end{array}$$

If, on the other hand, $j_i = 2$, and consequently $k = 1$, we have:

$$\begin{array}{rcll} p(o_i|a_2)v^1_2 & = & p(o_i|a_2)(v_2 + c) & \text{by definition of } \vec{v}^1 \\ & \geq & p(o_i|a_2)v_2 & \text{since } c \geq 0 \\ & = & p(o_i|a_1)v_1 & \text{since } j_i = 2 \\ & \geq & p(o_i|a_1)(v_1 - c) & \text{since } c \geq 0 \\ & = & p(o_i|a_1)v^1_1 & \text{by definition of } \vec{v}^1 \end{array}$$

The proof that for each i and $k \neq j_i$ the inequation $p(o_i|a_{j_i})v^1_{j_i} \geq p(o_i|a_k)v^1_k$ holds is analogous.

Hence we have proved that

$$P_e(\vec{v}^1) = \sum_i p(o_i|a_{j_i})v^1_{j_i} \quad \text{and} \quad P_e(\vec{v}^2) = \sum_i p(o_i|a_{j_i})v^2_{j_i}$$

and a simple calculation shows that (8) holds.

Inductive case Let $\vec{v} \in D^{(n)}$. Let c be the minimum between v_{n-1} and the minimum $x \geq 0$ such that for some i and k

$$\begin{array}{rcll} p(o_i|a_{n-1})(v_{n-1} - x) & = & p(o_i|a_n)(v_n + x) & \\ & \text{or} & & \\ p(o_i|a_{n-1})(v_{n-1} - x) & = & p(o_i|a_k)v_k & k \neq n \\ & \text{or} & & \\ p(o_i|a_{j_i})v_{j_i} & = & p(o_i|a_n)(v_n + x) & j_i \neq n-1 \end{array}$$

Analogously, let d be the minimum between v_{n+1} and the minimum $x \geq 0$ such that for some i and k

$$\begin{array}{rcll} p(o_i|a_n)(v_n - x) & = & p(o_i|a_{n-1})(v_{n-1} + x) & \\ & \text{or} & & \\ p(o_i|a_n)(v_n - x) & = & p(o_i|a_k)v_k & k \neq n-1 \\ & \text{or} & & \\ p(o_i|a_{j_i})v_{j_i} & = & p(o_i|a_{n-1})(v_{n-1} + x) & j_i \neq n \end{array}$$

Similarly to the base case, define $\vec{v}^1$, $\vec{v}^2$ as

$$\vec{v}^1 = (v_1, v_2, \ldots, v_{n-2}, v_{n-1} - c, v_n + c)$$

and

$$\vec{v}^2 = (v_1, v_2, \ldots, v_{n-2}, v_{n-1} + d, v_n - d)$$

and consider the same convex coefficients

$$\lambda_1 = \frac{d}{c+d} \qquad \lambda_2 = \frac{c}{c+d}$$

Again, we have $\vec{v} = \lambda_1 \vec{v}^1 + \lambda_2 \vec{v}^2$.

By case analysis, and following the analogous proof given for $n = 2$, we can prove that for each i and k the inequations $p(o_i|a_{j_i})v^1_{j_i} \geq p(o_i|a_k)v^1_k$ and $p(o_i|a_{j_i})v^2_{j_i} \geq p(o_i|a_k)v^2_k$ hold, hence, following the same lines as in the base case, we derive

$$P_e(\vec{v}) = \lambda_1 P_e(\vec{v}^1) + \lambda_2 P_e(\vec{v}^2)$$

We now prove that $\vec{v}^1$ and $\vec{v}^2$ can be obtained as convex combinations of corner points of P_e in the hyperplanes (instances of $D^{(n-1)}$) defined by the equations that give, respectively, the c and d above. More precisely, if $c = v_{n-1}$ the equation is $x_{n-1} = 0$. Otherwise, the equation is of the form

$$p(o_i|a_k)x_k = p(o_i|a_\ell)x_\ell$$

and analogously for d. We develop the proof for $\vec{v}^2$; the case of $\vec{v}^1$ is analogous.

If $d = 0$, then the hyperplane is defined by the equation $x_n = 0$, and it consists of the set of vectors of the form $(x_1, x_2, \ldots, x_{n-1})$. The Bayes' risk is defined in this

hyperplane exactly in the same way as P_e (since the contribution of x_n is null) and therefore the corner points are the same. By inductive hypothesis, those corner points are given by the solutions to the set of disequations of the form given in Definition 4.1. To obtain the corner points in $D^{(n)}$ it is sufficient to add the equation $x_n = 0$.

Assume now that d is given by one of the other equations. Let us consider the first one, the cases of the other two are analogous. Let us consider, therefore, the hyperplane $\mathcal{H}$ (instance of $D(n-1)$) defined by the equation

$$p(o_i|a_n)x_n = p(o_i|a_{n-1})x_{n-1} \tag{9}$$

It is convenient to perform a transformation of coordinates. Namely, represent the elements of $\mathcal{H}$ as vectors $\vec{y}$ with

$$y_j = \begin{cases} x_j & 1 \leq j \leq n-2 \\ x_{n-1} & j = n-1 \end{cases} \tag{10}$$

Consider the channel

$$\mathcal{C}' = \langle \mathcal{A}', \mathcal{O}, p'(\cdot|\cdot) \rangle$$

with $\mathcal{A}' = \{a_1, a_2, \ldots, a_{n-1}\}$, and

$$p'(o_k|a_j) = \begin{cases} p(o_k|a_j) & 1 \leq j \leq n-2 \\ \max\{p_1(k), p_2(k)\} & j = n-1 \end{cases}$$

where

$$p_1(k) = p(o_k, a_{n-1}) \frac{p(o_i|a_n)}{p(o_i|a_{n-1}) + p(o_i|a_n)}$$

and

$$p_2(k) = p(o_k, a_n) \frac{p(o_i|a_{n-1})}{p(o_i|a_{n-1}) + p(o_i|a_n)}$$

The Bayes's risk in $\mathcal{H}$ is defined by

$$P_e(\vec{y}) = \sum_k \max_{1 \leq j \leq n-1} p'(o_k|a_j) y_j$$

and a simple calculation shows that $P_e(\vec{y}) = P_e(\vec{x})$ whenever $\vec{x}$ satisfies (9) and $\vec{y}$ and $\vec{x}$ are related by (10). Hence the corner points of $P_e(\vec{x})$ over $\mathcal{H}$ can be obtained from those of $P_e(\vec{y})$.

The systems of inequations in $\mathbb{S}(\mathcal{C})$ are obtained from those in $\mathbb{S}(\mathcal{C}')$ in the following way. For each system in $\mathbb{S}(\mathcal{C}')$, replace the equation $y_1 + y_2 + \ldots + y_{n-1} = 1$ by $x_1 + x_2 + \ldots + x_{n-1} + x_n = 1$, and replace, in each equation, every occurrence of y_j by x_j, for j from 1 to $n-2$. Furthermore, if y_{n-1} occurs in an equation E of the form $y_{n-1} = 0$, then replace E by the equations $x_{n-1} = 0$ and $x_n = 0$. Otherwise, it must be the case that for some k, $p'(o_k|a_{n-1})y_{n-1}$ occurs in some (two) of the other equations. In that case, replace that expression by $p(o_k|a_{n-1})x_{n-1}$ if $p_1(k) \geq p_2(k)$, and by $p(o_k|a_n)x_n$ otherwise. The transformation to apply on the inequational part is trivial. □

Note that $\mathbb{S}(\mathcal{C})$ is finite, hence the U in Theorem 4.2 is finite as well.

Example 4.3 (Binary hypothesis testing) *The case $n = 2$ is particularly simple: the systems generated by $\mathcal{C}$ are all those of the form*

$$\{p(o_i|a_1)x_1 = p(o_i|a_2)x_2 \ , \ x_1 + x_2 = 1\}$$

plus the two systems

$$\{x_1 = 0 \ , \ x_1 + x_2 = 1\}$$
$$\{x_2 = 0 \ , \ x_1 + x_2 = 1\}$$

These systems are always solvable, hence we have $m + 2$ corner points, where we recall that m is the cardinality of O.

Let us illustrate this case with a concrete example: let $\mathcal{C}$ be the channel determined by the following matrix:

	o_1	o_2	o_3
a_1	$1/2$	$1/3$	$1/6$
a_2	$1/6$	$1/2$	$1/3$

The systems generated by $\mathcal{C}$ are:

$$\{x_1 = 0 \quad , \quad x_1 + x_2 = 1\}$$
$$\{\tfrac{1}{2}x_1 = \tfrac{1}{6}x_2 \quad , \quad x_1 + x_2 = 1\}$$
$$\{\tfrac{1}{3}x_1 = \tfrac{1}{2}x_2 \quad , \quad x_1 + x_2 = 1\}$$
$$\{\tfrac{1}{6}x_1 = \tfrac{1}{3}x_2 \quad , \quad x_1 + x_2 = 1\}$$
$$\{x_1 = 0 \quad , \quad x_1 + x_2 = 1\}$$

The solutions of these systems are: $(0, 1)$, $(1/4, 3/4)$, $(3/5, 2/5)$, $(2/3, 1/3)$, and $(1, 0)$, respectively. The value of P_e on these points is 0, $1/4$, $3/10$ (maximum), $1/3$, and 0 respectively, and P_e is piecewise linear between these points, i.e. it can be generated by convex combination of these points and its value on them. Its graph is illustrated in Figure 1, where x_1 is represented by x and x_2 by $1 - x$.

Example 4.4 (Ternary hypothesis testing) *Let us consider now a channel $\mathcal{C}$ with three inputs. Assume the channel has the following matrix:*

	o_1	o_2	o_3
a_1	$2/3$	$1/6$	$1/6$
a_2	$1/8$	$3/4$	$1/8$
a_3	$1/10$	$1/10$	$4/5$

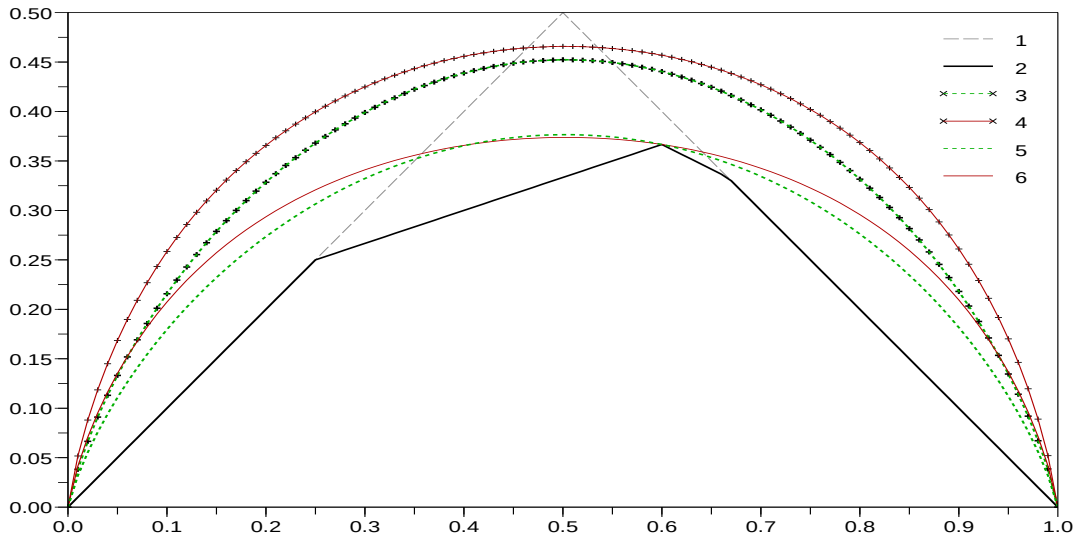

Figure 1. The graph of the Bayes' risk for the channel in Example 4.3 and various bounds for it. Curve 1 represents the probability of error if we ignore the observables, i.e. the function $f(\vec{x}) = 1 - \max_j x_j$. Curve 2 represents the Bayes' risk $P_e(\vec{x})$. Curve 3 represents the Hellman-Raviv bound $\frac{1}{2}H(A|O)$. Curve 4 represents the Santhi-Vardy bound $1 - 2^{-H(A|O)}$. Finally, Curves 5 and 6 represent the improvements on 3 and 4, respectively, that we get by applying the method induced by our Proposition 3.5.

The following is an example of a solvable system generated by $\mathcal{C}$:

$$\begin{aligned} \tfrac{2}{3}x_1 &= \tfrac{1}{8}x_2 \\ \tfrac{1}{8}x_2 &= \tfrac{4}{5}x_3 \\ x_1 + x_2 + x_3 &= 1 \\ \tfrac{2}{3}x_1 &\geq \tfrac{1}{10}x_3 \\ \tfrac{1}{8}x_2 &\geq \tfrac{1}{6}x_1 \end{aligned}$$

Another example is

$$\begin{aligned} \tfrac{1}{6}x_1 &= \tfrac{3}{4}x_2 \\ x_3 &= 0 \\ x_1 + x_2 + x_3 &= 1 \end{aligned}$$

The graph of P_e is depicted in Figure 2, where x_3 is represented by $1 - x_1 - x_2$.

5 Maximum Bayes' risk and relation with strong anonymity

In this section we discuss the Bayes' risk in the extreme cases of maximum and minimum (i.e. 0) capacity, and, in the second case, we illustrate the relation with the notion of probabilistic strong anonymity existing in literature.

Maximum capacity If the channel has no noise, which means that for each observable o there exists at most one a such that $p(o|a) \neq 0$, then the Bayes' risk is 0 for every input's distribution. In fact

$$\begin{aligned} P_e(\vec{x}) &= 1 - \textstyle\sum_o \max_j p(o|a_j)x_j \\ &= 1 - \textstyle\sum_j \sum_o p(o|a_j)x_j \\ &= 1 - \textstyle\sum_j x_j = 0 \end{aligned}$$

Capacity 0 The case in which the capacity of the channel is 0 is by definition obtained when $I(A;O) = 0$ for all possible input distributions of $\mathcal{A}$. From information theory we know that this is the case iff A and O are independent (cfr. [8], page 27). Hence we have the following characterization:

Proposition 5.1 *Given an anonymity system $\langle \mathcal{A}, \mathcal{O}, p(\cdot|\cdot)\rangle$, the capacity of the corresponding channel is 0 iff all the rows of the channel matrix are the same, i.e. $p(o|a) = p(o|a')$ for all o, a, a'.*

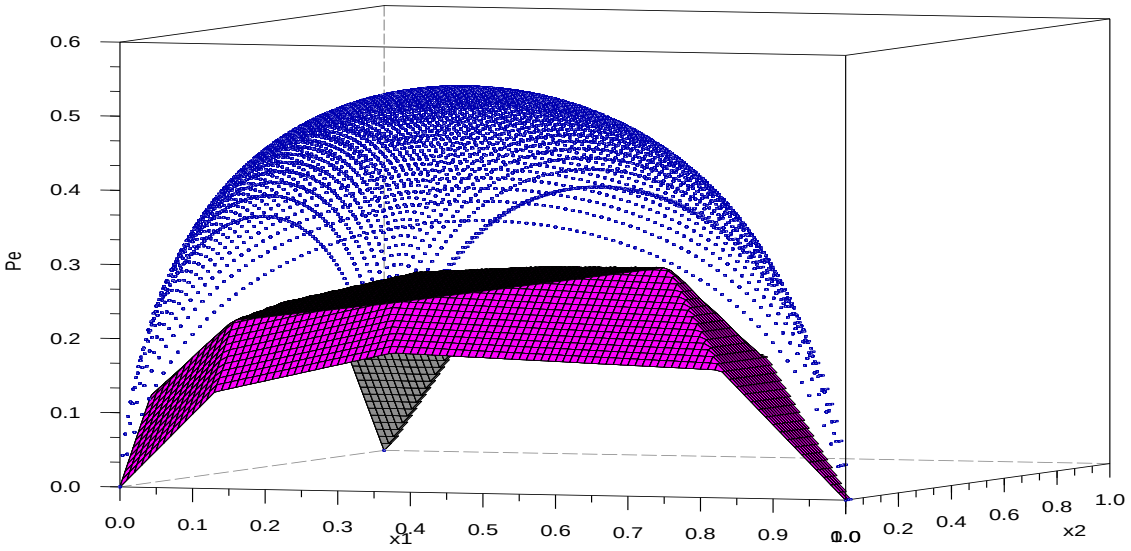

Figure 2. Ternary hypothesis testing. The solid curve represents the Bayes' risk for the channel in Example 4.4, while the dotted curve represents the Santhi-Vardy bound $1 - 2^{-H(A|O)}$.

The condition $p(o|a) = p(o|a')$ *for all* o, a, a' has been called *strong probabilistic anonymity* in [1] and it is equivalent to the condition $p(a|o) = p(a)$ *for all* o, a. The latter was considered as a definition of anonymity in [4] and it is called *conditional anonymity* in [12].

Capacity 0 is the optimal case also w.r.t. the capability of the adversary of inferring the hidden information. In fact, we can prove that the Bayes' risk achieves its highest possible value, for a given n (cardinality of $\mathcal{A}$), when the rows of the matrix are all the same and the distribution is uniform. In this case, we have

$$\begin{aligned} P_e(\tfrac{1}{n}, \tfrac{1}{n}, \ldots, \tfrac{1}{n}) &= 1 - \textstyle\sum_o \max_j p(o|a_j) x_j \\ &= 1 - \textstyle\sum_o p(o|a) \frac{1}{n} \\ &= 1 - \tfrac{1}{n} \textstyle\sum_o p(o|a) \\ &= \tfrac{n-1}{n} \end{aligned}$$

An example of protocol with capacity 0 is the *dining cryptographers* in a connected graph [4], under the assumption that it is always one of the cryptographers who pays, and that the coins are fair.

6 Application: Crowds

In this section we discuss how to compute the channel matrix for a given protocol using automated tools, and use it to improve the bound for the probability of error. We illustrate our ideas on a variation of Crowds, a well-known anonymity protocol from the literature.

In this protocol, introduced by Reiter and Rubin in [21], a user (called the *initiator*) wants to send a message to a web server without revealing its identity. To achieve that, he routes the message through a crowd of users participating in the protocol. The routing is performed using the following protocol: in the beginning, the initiator selects randomly a user (called a *forwarder*), possibly himself, and forwards the request to him. A forwarder, upon receiving a message, performs a probabilistic choice. With probability p_f (a parameter of the protocol) he selects a new user and forwards once again the message. With probability $1 - p_f$ he sends the message directly to the server.

It is easy to see that the initiator is strongly anonymous wrt the server, as all users have the same probability of being the forwarder who finally delivers the message. However, the more interesting case is when the attacker is one of the users of the protocol (called a *corrupted* user) which uses his information to find out the identity of the initiator. A corrupted user has more information than the server since he sees other users forwarding the message through him. The initiator, being the in first in the path, has greater probability of forwarding the message to the attacker than any other user, so strong anonymity cannot hold. However, under certain conditions on the number of corrupted

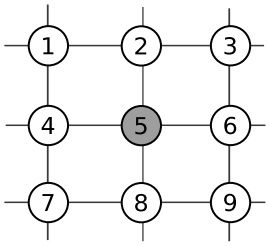

Figure 3. An instance of Crowds with nine users in a grid network. User 5 is the only corrupted one.

	d_2	d_4	d_6	d_8
u_1	0.33	0.33	0.17	0.17
u_3	0.33	0.17	0.33	0.17
u_7	0.17	0.33	0.17	0.33
u_9	0.17	0.17	0.33	0.33
u_2	0.68	0.07	0.07	0.17
u_4	0.07	0.68	0.17	0.07
u_6	0.07	0.17	0.68	0.07
u_8	0.17	0.07	0.07	0.68

Figure 4. The channel matrix of the examined instance of Crowds. The symbols u_i, d_j mean that user i is the initiator and user j was detected respectively.

users, Crowds can be shown to satisfy a weaker notion of anonymity called *probable innocence*.

In the original protocol, all users are considered to be able to communicate with any other user, in other words the connection graph is a clique. To make the example more interesting, we consider a more restricted grid-shaped network as shown in Figure 3. In this network there is a total of nine users, each of whom can only communicate with the four that are adjacent to him. We assume that the network "wraps" at the edges, so user 1 can communicate with both user 3 and user 7. Also, we assume that the only corrupted user is user 5.

To construct the channel matrix of the protocol, we start by identifying the set of anonymous facts, which depends on what the system is trying to hide. In protocols where one user performs an action of interest (like initiating a message in our example) and we want to protect his identity, the set $\mathcal{A}$ would be the set of the users of the protocol. Note that the corrupted users should not be included in this set, since we cannot expect the attacker's own actions to be hidden from him. So in our case we have $\mathcal{A} = \{u_1, u_2, u_3, u_4, u_6, u_7, u_8, u_9\}$ where u_i means that user i is the initiator.

The set of observables should also be defined, based on the visible actions of the protocol and on the various assumptions made about the attacker. In Crowds we assume that the attacker does not have access to the entire network (such an attacker would be too powerful for this protocol) but only to the messages that pass through a corrupted user. Each time that a user i forwards the message to a corrupted user we say that he is *detected* which corresponds to an observable action in the protocol. Along the lines of other studies of Crowds (eg [25]) we consider that an attacker will not forward a message himself, since by doing so he would not gain more information. So at each execution there is at most one detected user and since only the users 2, 4, 6 and 8 can communicate with the corrupted user, we have $\mathcal{O} = \{d_2, d_4, d_6, d_8\}$ where d_j means that user j was detected. As we explain later, there is also a non-zero probability that no user is detected, which is the case when the message arrives to the server without passing by user 5.

After defining $\mathcal{A}, \mathcal{O}$ we should model the protocol in some formal probabilistic language. In our example, we have modeled Crowds in the language of the PRISM model-checker, that is essentially a formalism to describe Markov Decision Processes. Then the channel matrix of conditional probabilities $p(o|a)$ must be computed, either by hand or by using an automated tool like PRISM which can compute the probability of reaching a specific state starting from a given one. Thus, each conditional probability $p(d_j|u_i)$ is computed as the probability of reaching a state where the attacker has detected user j, starting from the state where i is the initiator. If $p_f < 1$ there is always a non-zero probability of not detecting any user at all, which happens if the message arrives at the server without passing by user 5. In this case, the execution of the protocol passes completely unnoticed by the adversary. Thus, in our analysis, we compute all probabilities conditioned on the fact that *some* observation was made. This corresponds to normalizing the rows of the table, that is dividing all $p(o|a_i)$ by $\sum_i p(o|a_i)$.

In Figure 4 the channel matrix is displayed for the examined Crowds instance, computed using a probability of forwarding $p_f = 0.8$. We have split the users in two groups, the ones who cannot communicate directly with the corrupted user, and the ones who can. When a user of the first group, say user 1, is the initiator, there is a higher probability of detecting the users that are adjacent to him (users 2 and 4) than the other two (users 6 and 8) since the message needs two steps to arrive to the latters. So $p(d_2|u_1) = p(d_4|u_1) = 0.33$ are greater than $p(d_6|u_1) = p(d_8|u_1) = 0.17$. In the second group users have direct communication to the attacker, so when user 2 is the initiator, the probability $p(d_2|u_2)$ of detecting him is high. From the remaining three observables d_8 has higher probability since user 8 can be reached from user 2 in one

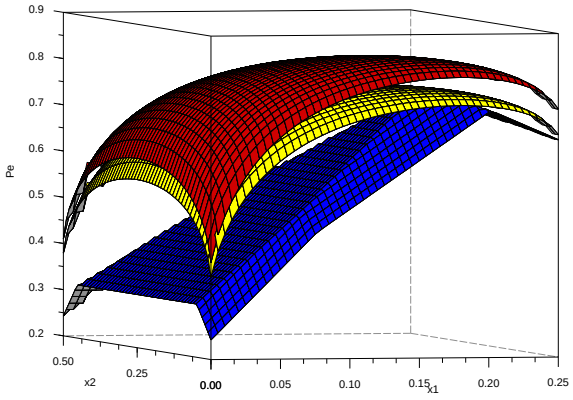

Figure 5. The lower curve is the probability of error in the examined instance of Crowds. The upper two are the Santhi and Vardy's bound and it's improved version.

step, while users 4 and 6 need two steps. Inside each group the rows are symmetric since the users behave similarly. However between the groups the rows are different which is caused by the different connectivity to the corrupted user 5.

We can now compute the probability of error for this instance of Crowds, which is displayed in the lower curve of Figure 5. Since we have eight users, to plot this function we have to map it to the three dimensions. We do this by considering the users 1, 3, 7, 9 to have the same probability x_1, the users 2, 8 to have the same probability x_2 and the users 4, 6 to have the same probability $1-x_1-x_2$. Then we plot P_e as a function of x_1, x_2 in the ranges $0 \leq x_1 \leq 1/4$, $0 \leq x_2 \leq 1/2$. Note that when $x_1 = x_2 = 0$ there are still two users (4, 6) among whom the probability is distributed, so P_e is not 0. The upper curve of Figure 5 shows the Santhi and Vardy's bound on the probability of error. Since all the rows of the matrix are different the bound is not a tight one as it can be seen in the Figure.

We can obtain a better bound by applying Proposition 3.5. The set of corner points, characterized by Theorem 4.2, is finite and can be automatically constructed by solving the corresponding systems of inequations. After computing the corner points, it is sufficient to take $c = \max_u P_e(\vec{u})/h(\vec{u})$, where h is the original bound, and take $c\,h$ as the improved bound. In our example we found $c = 0.925$ which was given for the corner point $\vec{u} = (0.17, 0.17, 0.17, 0.17, 0.08, 0.08, 0.08, 0.08)$.

References

[1] M. Bhargava and C. Palamidessi. Probabilistic anonymity. In M. Abadi and L. de Alfaro, editors, *Proceedings of CONCUR*, volume 3653 of *Lecture Notes in Computer Science*, pages 171–185. Springer, 2005. Available at http://www.lix.polytechnique.fr/∼catuscia/papers/Anonymity/concur.pdf.

[2] K. Chatzikokolakis and C. Palamidessi. Probable innocence revisited. *Theoretical Computer Science*, 367(1-2):123–138, 2006. Available at http://www.lix.polytechnique.fr/∼catuscia/papers/Anonymity/tcsPI.pdf.

[3] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. Anonymity protocols as noisy channels. In *Postproceedings of the Symp. on Trustworthy Global Computing*, Lecture Notes in Computer Science. Springer, 2006. To appear. Available at http://www.lix.polytechnique.fr/∼catuscia/papers/Anonymity/Channels/full.pdf.

[4] D. Chaum. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, 1:65–75, 1988.

[5] D. Clark, S. Hunt, and P. Malacaria. Quantitative analysis of the leakage of confidential data. In *Proc. of QAPL 2001*, volume 59 (3) of *Electr. Notes Theor. Comput. Sci*, pages 238–251. Elsevier Science B.V., 2001.

[6] D. Clark, S. Hunt, and P. Malacaria. Quantified interference for a while language. In *Proc. of QAPL 2004*, volume 112 of *Electr. Notes Theor. Comput. Sci*, pages 149–166. Elsevier Science B.V., 2005.

[7] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Designing Privacy Enhancing Technologies, International Workshop on Design Issues in Anonymity and Unobservability*, volume 2009 of *Lecture Notes in Computer Science*, pages 44–66. Springer, 2000.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[9] Y. Deng, J. Pang, and P. Wu. Measuring anonymity with relative entropy. In *Proceedings of the 4th International Workshop on Formal Aspects in Security and Trust (FAST)*, Lecture Notes in Computer Science. Springer, 2006. To appear.

[10] C. Díaz, S. Seys, J. Claessens, and B. Preneel. Towards measuring anonymity. In R. Dingledine and P. F. Syverson, editors, *Proceedings of the workshop on Privacy Enhancing Technologies (PET) 2002*, volume 2482 of *Lecture Notes in Computer Science*, pages 54–68. Springer, 2002.

[11] J. W. Gray, III. Toward a mathematical foundation for information flow security. In *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy (SSP '91)*, pages 21–35, Washington - Brussels - Tokyo, May 1991. IEEE.

[12] J. Y. Halpern and K. R. O'Neill. Anonymity and information hiding in multiagent systems. *Journal of Computer Security*, 13(3):483–512, 2005.

[13] M. Hellman and J. Raviv. Probability of error, equivocation, and the chernoff bound. *IEEE Trans. on Information Theory*, IT–16:368–372, 1970.

[14] G. Lowe. Quantifying information flow. In *Proc. of CSFW 2002*, pages 18–31. IEEE Computer Society Press, 2002.

[15] U. M. Maurer. Authentication theory and hypothesis testing. *IEEE Transactions on Information Theory*, 46(4):1350–1356, 2000.

[16] J. McLean. Security models and information flow. In *IEEE Symposium on Security and Privacy*, pages 180–189, 1990.

[17] I. S. Moskowitz, R. E. Newman, D. P. Crepeau, and A. R. Miller. Covert channels and anonymizing networks. In S. Jajodia, P. Samarati, and P. F. Syverson, editors, *WPES*, pages 79–88. ACM, 2003.

[18] I. S. Moskowitz, R. E. Newman, and P. F. Syverson. Quasi-anonymous channels. In *IASTED CNIS*, pages 126–131, 2003.

[19] A. D. Pierro, C. Hankin, and H. Wiklicky. Approximate non-interference. *Journal of Computer Security*, 12(1):37–82, 2004.

[20] A. D. Pierro, C. Hankin, and H. Wiklicky. Measuring the confinement of probabilistic systems. *Theoretical Computer Science*, 340(1):3–56, 2005.

[21] M. K. Reiter and A. D. Rubin. Crowds: anonymity for Web transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, 1998.

[22] A. Rényi. On the amount of missing information and the Neyman-Pearson lemma. In *Festschriftf for J. Neyman*, pages 281–288. Wiley, New York, 1966.

[23] N. Santhi and A. Vardy. On an improvement over Rényi's equivocation bound, 2006. Presented at the 44-th Annual Allerton Conference on Communication, Control, and Computing, September 2006. Available at http://arxiv.org/abs/cs/0608087.

[24] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In R. Dingledine and P. F. Syverson, editors, *Proceedings of the workshop on Privacy Enhancing Technologies (PET) 2002*, volume 2482 of *Lecture Notes in Computer Science*, pages 41–53. Springer, 2002.

[25] V. Shmatikov. Probabilistic model checking of an anonymity system. *Journal of Computer Security*, 12(3/4):355–377, 2004.

[26] P. Syverson, D. Goldschlag, and M. Reed. Anonymous connections and onion routing. In *IEEE Symposium on Security and Privacy*, pages 44–54, Oakland, California, 1997.

7 Appendix

We give here the proof of Proposition 3.6.

Proposition 3.6 *Let $f : D^{(n)} \to \mathbb{R}$ be the function $f(\vec{x}) = 1 - \max_j x_j$. Then f is convexly generated by $f(U)$ with $U = U_1 \cup U_2 \cup \ldots \cup U_n$ where, for each k, U_k is the set of all vectors that have value $1/k$ in exactly k components, and 0 everywhere else.*

Proof Observe that f coincides with the Bayes' risk for a channel $\mathcal{C}$ with 0 capacity, i.e. a channel in which for every o, a, a' we have $p(o|a) = p(o|a')$. In fact, the Bayes's risk for such channel is given by

$$\begin{array}{rcll} P_e(\vec{x}) & = & 1 - \sum_o \max_j p(o|a_j)x_j & \\ & = & 1 - \sum_o p(o|a) \max_j x_j & \text{for a choosen } a \\ & = & 1 - \max_j x_j & \text{since } \sum_o p(o|a) = 1 \end{array}$$

By Theorem 4.2, P_e is convexly generated by $P_e(U)$, where U is the set of solutions of the solvable systems in $\mathbb{S}(\mathcal{C})$. Now, each such system is of the form

$$\begin{array}{rcl} p(o_{i_1}|a_{j_1})x_{j_1} & = & p(o_{i_1}|a_{j_2})x_{j_2} \\ p(o_{i_2}|a_{j_2})x_{j_2} & = & p(o_{i_2}|a_{j_3})x_{j_3} \\ & \vdots & \\ p(o_{i_r}|a_{j_{k-1}})x_{j_{k-1}} & = & p(o_{i_r}|a_{j_k})x_{j_k} \\ x_j & = & 0 \quad \text{for } j \notin \{j_1, j_2, \ldots, j_k\} \\ x_1 + x_2 + \ldots + x_n & = & 1 \\ p(o_{i_h}|a_{j_h})x_{j_h} & \geq & p(o_{i_h}|a_\ell)x_\ell \quad \text{for } 1 \leq h, \ell \leq k, n \end{array}$$

Which, given the fact that for all i, j, j' the equality $p(o_i|a_j) = p(o_i|a_{j'})$ holds, can be simplified to

$$\begin{array}{rcl} x_{j_1} & = & x_{j_2} \\ x_{j_2} & = & x_{j_3} \\ & \vdots & \\ x_{j_{k-1}} & = & x_{j_k} \\ x_j & = & 0 \quad \text{for } j \notin \{j_1, j_2, \ldots, j_k\} \\ x_1 + x_2 + \ldots + x_n & = & 1 \\ x_{j_h} & \geq & x_\ell \quad \text{for } 1 \leq h, \ell \leq k, n \end{array}$$

A simple calculation shows that such a system has one (and only one) solution $\vec{u} = (u_1, u_2, \ldots, u_n)$ where

$$u_j = \begin{cases} \frac{1}{k} & \text{if } j \in \{j_1, j_2, \ldots, j_k\} \\ 0 & \text{otherwise} \end{cases}$$

which concludes the proof. □